John Raven

Final Project

The data set I chose is the number of wins by the National Football League franchise the Baltimore Ravens. With this data set I am looking specifically at the 1996 through the 2020 seasons, this is due to a 17th game being added to the regular season during the 2021 seasons. There has been a total of 400 games from the 1996 season through the 2020 season, the Ravens have won a total of 225 games, lost 175 games, and tied 1 game. For the purpose of the tests, I will be conducting I have moved the tie game to the loss sample set because I will be looking specifically at wins.

The first set of tests I ran were mean, median, mode, standard deviation, and variance. The first question that I asked was “What is the average number of wins per season for the Ravens franchise?” The output that would correlate to this question is the following: . The mean method will read through the list of wins, add them up and divide by the number of elements. In this case the number of elements is 25, due to the 25 seasons. The second question that I asked was “What is the median of wins per season for the Ravens franchise?” The output that correlates to this question is the following: . The median method sorts the list, reads through it, and then outputs the median of the list. The third question that I asked was “What is the mode of the list of wins for the Ravens franchise?” The output that correlates to this question is the following: . The mode method reads through the list, finds duplicate entries using a count and highCount variable, this method then outputs the highCount variable in the form of an integer. The next question that I asked was “What is the standard deviation of the list of wins per season for the Ravens franchise?” The output that corresponds to this question is the following: . This method uses the standard deviation equation and calculates the standard deviation of the list of wins. The last question I asked for this portion of my project was “What is the variance of the list of wins per season for the Ravens franchise?” The output that corresponds to this question is the following:. The variance method uses the variance formula and calculates the variance of the list of wins.

The second set of tests I ran were permutation, combination, union, intersection, and compliment against the data set. The question for the permutation test that I asked was “How many different ways can you order the Ravens’ seasons with more than 10 wins?” The output that correlates to this question is the following: . The permutation method calculates the different ways winning 10 or more wins can be ordered. The next question I asked for the combination test was “How many different ways can the Ravens win 10 games among 16 games?” The output that correlates to this question is the following: . The combination method calculates the different ways the Ravens can win 10 games in the 16-game regular season. The next question I asked was for the union test, “What is the union between the list of wins and its compliment?” The out that correlates to this question is the following: . The union method reads through the wins list and the compliment of the wins list and performs the union of the two lists. The next question that I asked was “What is the intersection between the list of wins and its compliment?” The output that correlates to this question is the following: . We can tell the empty set is the correct answer simply because the compliment is what the sample set does not contain, so there should be nothing in common between the sample set and its compliment. The final question that I asked for this section was “What is the compliment of the list of wins?” The output that correlates to this method is the following: . The compliment method reads the wins list and compares it to a preset list, in this case the preset list was the list of losses, it then compares the lists and will produce a list containing elements that the preset list has, and the wins list does not.

The third set of tests that I ran against my data set were binomial distribution, geometric distribution, hyper geometric distribution, Poisson distribution, Chebyshev’s theorem, and uniform distribution. The first question I asked for this section was “What are the odds of picking 15 random seasons that exceed 8 wins?” The output that corresponds to this question is the following: . The binDistro method uses the binomial distribution formula and takes in three variables, in this case those variables are numSeasons (25), 15 (y), and .6 (p). .6 represents the probability of picking a season that exceeds 8 wins. The next question that I asked for this section was “What are the odds of randomly picking a season that exceeds 8 wins by the 5th attempt?” The output that corresponds to this question is the following: . The geoDistro method takes in two variables and applies them to the geometric distribution formula. In this case the variables are 5 (r) and .6 (p). Once again, the .6 represents the probability of picking a season that exceeds 8 wins. The next question that I asked was “What are the odds of picking two seasons that have an odd amount of wins and exceed 10 wins?” The output that corresponds to this question is the following: . The hypGeoDistro method takes in two variables and puts them into the hyper geometric distribution formula. In this case it takes in an n variable which is 9, an r variable which is 3, an N variable which is numSeasons (25), and a y variable which is 2. The variable n is 9 because there are only 9 seasons with an odd number of wins; variable r is 3 because there are only 3 seasons that are odd and exceed 10 wins. The next question that I asked was “There are 16 games per season, on average the Ravens win 9 of them. What are the odds the Ravens win more than 8 games per season?” The output that corresponds to this question is the following: . The poisson method takes in two variables setEvents and event, in this case setEvents is gamesPerSeason and event is nine. The poisson method then applies these variables to the Poisson distribution. The next question that I asked was “What are the odds of the Ravens winning between 6 and 12 games per season with a mean of 9 and a standard deviation of 2.65?” The output that corresponds to this question is the following: . The cheby method takes in four variables, upper, lower, littleO, and mean. In this case the upper variable is 12, the lower variable is 6, littleO is 2.65, and the mean is 9, when calling the method, I also called my mean method instead of manually calculating the mean. The cheby method then takes these variables and applies them to Chebyshev’s theorem. The next question that I asked was “What are the odds of choosing the last 3 seasons out of the whole data set?” The output that corresponds to that is the following: . The uniDistro method takes in three variables, a, b, and x; in this case a is 1, b is numSeasons (25), and x is 3. The uniDistro method takes these three variables and applies them to the uniform distribution equation.

In conclusion I didn’t learn too much from this data set but, I did realize how difficult it can be to apply a single data set to the number of distributions and statistical analysis tools that we have learned. I am now able to understand why so many different equations and methods are necessary when analyzing data, and that not all data will fit all statistical analysis equations. I had a difficult time formulating necessary questions to fulfill these distributions properly at times. Through this project I was able to see the appeal to running some of these tests against data sets, especially sports teams. With statistical data sports teams would be able to easily see different trends in performance when analyzing the state of a team over different seasons.